



# **EOSDIS**

NASA'S EARTH OBSERVING SYSTEM  
DATA AND INFORMATION SYSTEM

# Making Metadata Better with the CMR and MMT

Summer ESIP 2016

Jason Gilman

Dana Shum

The material is based upon work supported by the National Aeronautics and Space  
Administration under Contract Number **NNG15HZ39C**

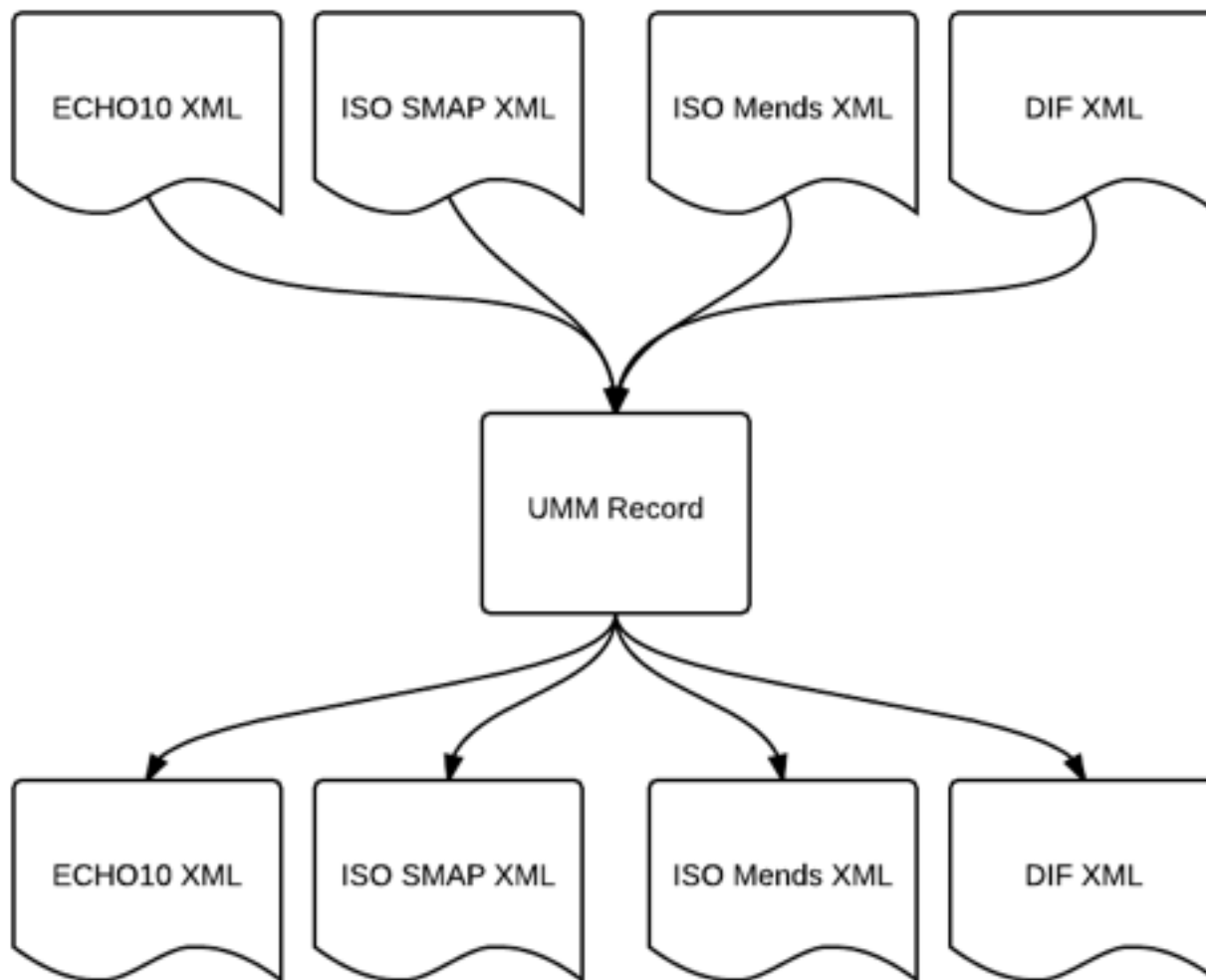
# **METADATA PROBLEMS**

# Completeness and Consistency

- Misspellings: “Bioosphere”
- Legacy Terms: “AM-1” instead of Terra
- Inconsistent Names: Processing levels “Level 1”, “1”
- Whitespace around element values
- Missing elements

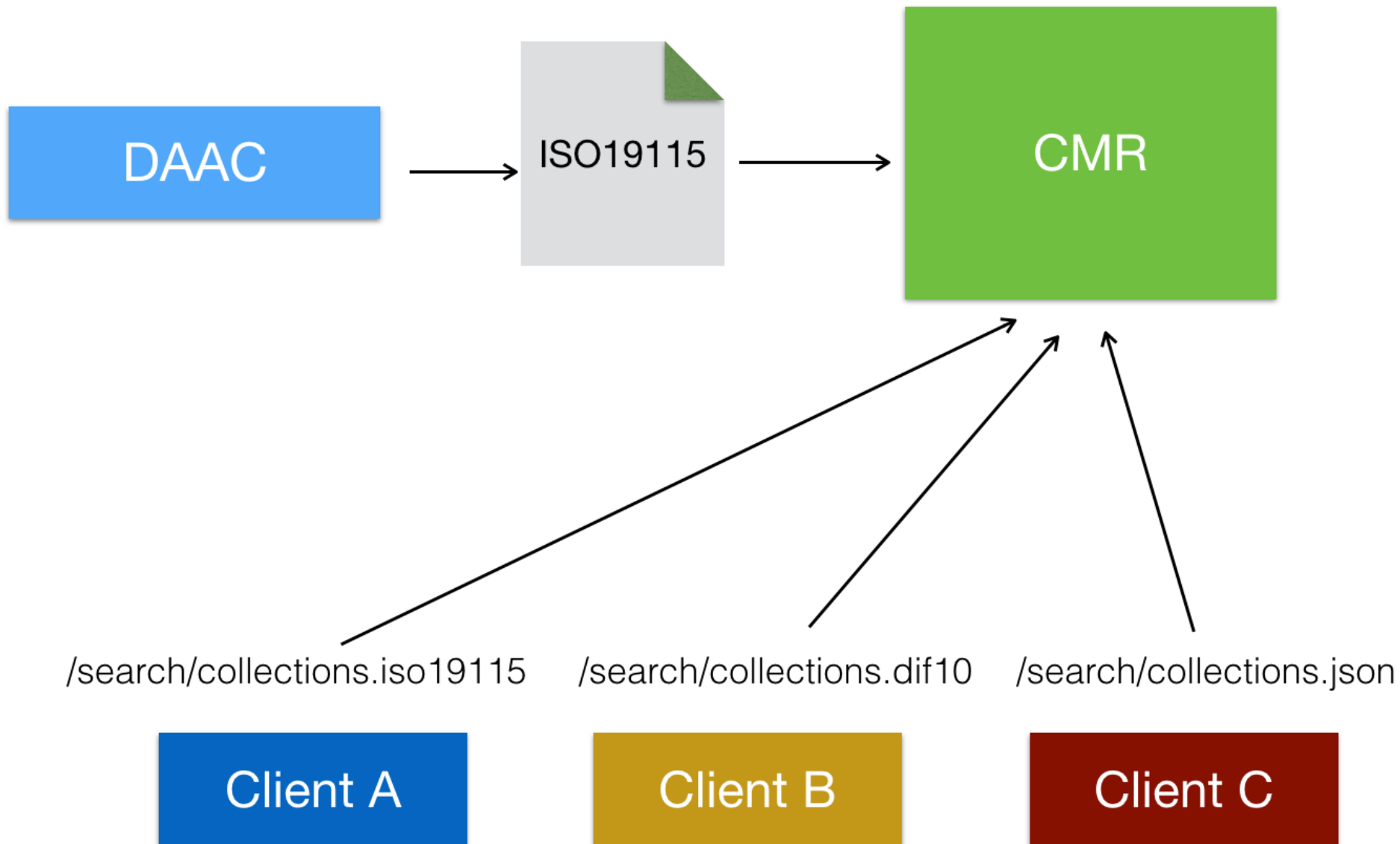
Overview and Benefits

# UNIFIED METADATA MODEL

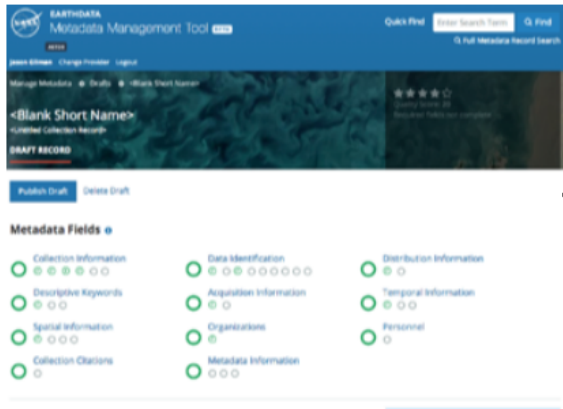


# UMM Benefits

- Common validations across all metadata dialects
- Consistent features across all dialects.
- Allows conversion from any dialect to any other dialect.



# MMT



{UMM  
JSON}

## CMR

/search/collections.iso19115

/search/collections.dif10

/search/collections.json

Client A

Client B

Client C

# **COMPLETENESS AND CONSISTENCY ENABLERS**

# MMT Pre-Ingest Validation

The screenshot displays the 'testRecord1\_001' draft record in the MMT system. The interface includes a top navigation bar with 'Manage Metadata', 'Drafts', and 'testRecord1\_001'. A sidebar on the right shows 'VERSION 001', a star rating, and a 'Quality Score: 20' with a note 'Required fields not complete'. The main content area features a 'DRAFT RECORD' header, 'Publish Draft' and 'Delete Draft' buttons, and a 'Metadata Fields' section. The 'Metadata Fields' section lists several categories with progress indicators: 'Collection Information' (4 required, 2 optional), 'Descriptive Keywords' (1 required, 2 optional), 'Spatial Information' (1 required, 3 optional), 'Collection Citations' (1 optional), 'Organizations' (1 required, 1 optional, 1 does not pass validation), 'Personnel' (1 required, 1 optional), and 'Metadata Information' (1 required, 2 optional, 1 optional). A modal dialog box is overlaid on the screen, providing a legend for the icons: a green circle with a white 'r' for 'Required but not complete', a green circle with a white checkmark for 'Required and complete', a white circle for 'Optional and not complete', a grey circle for 'Optional and complete', and a red circle with a white minus sign for 'Does not pass validation'. The dialog also has a 'Close' button.

Manage Metadata Drafts testRecord1\_001

testRecord1\_001  
This is test record 1

**DRAFT RECORD**

Publish Draft Delete Draft

**Metadata Fields**

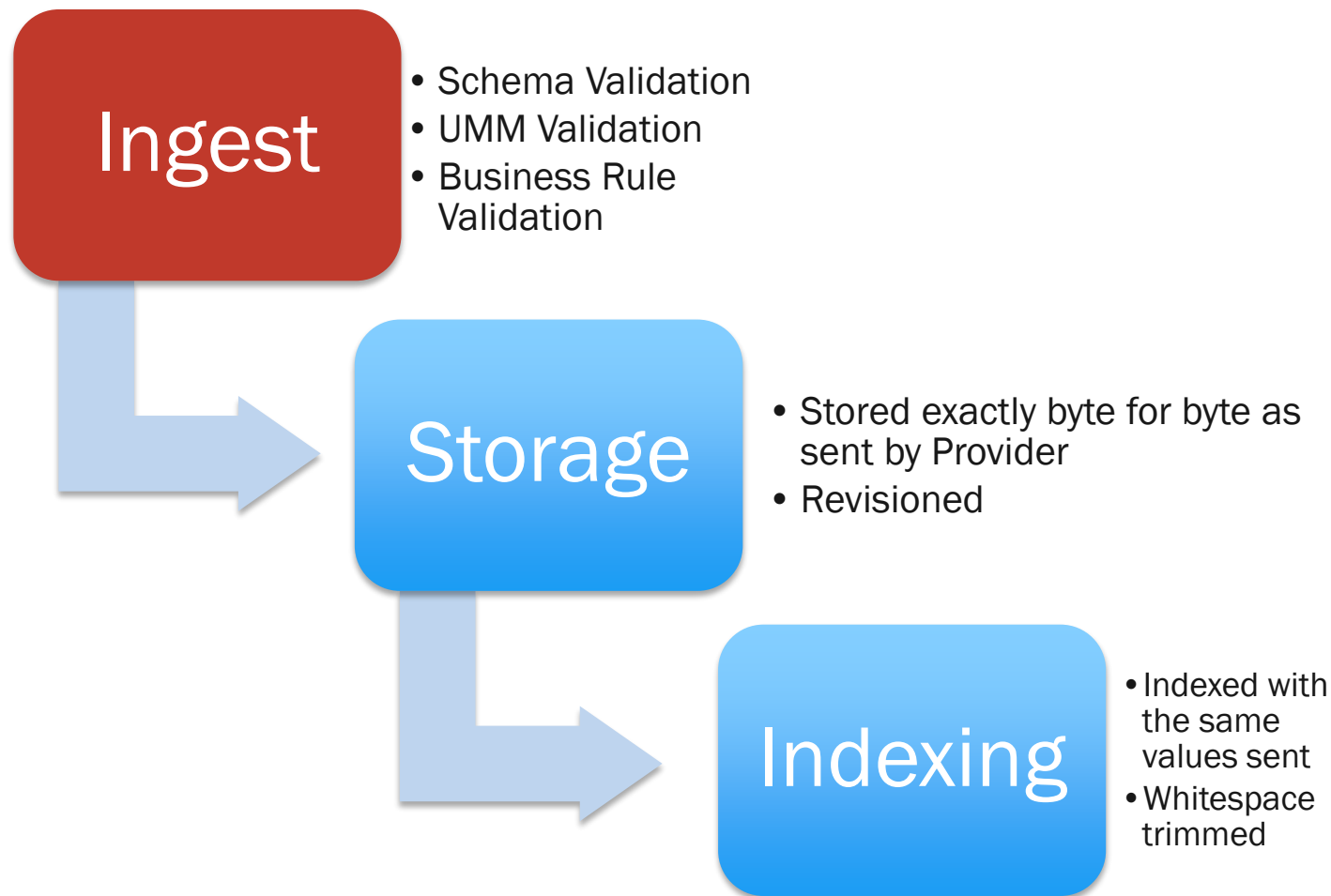
- Collection Information
- Descriptive Keywords
- Spatial Information
- Collection Citations
- Organizations
- Personnel
- Metadata Information

The icons below each form name indicate progress toward completion. See below for information on what each icon represents.

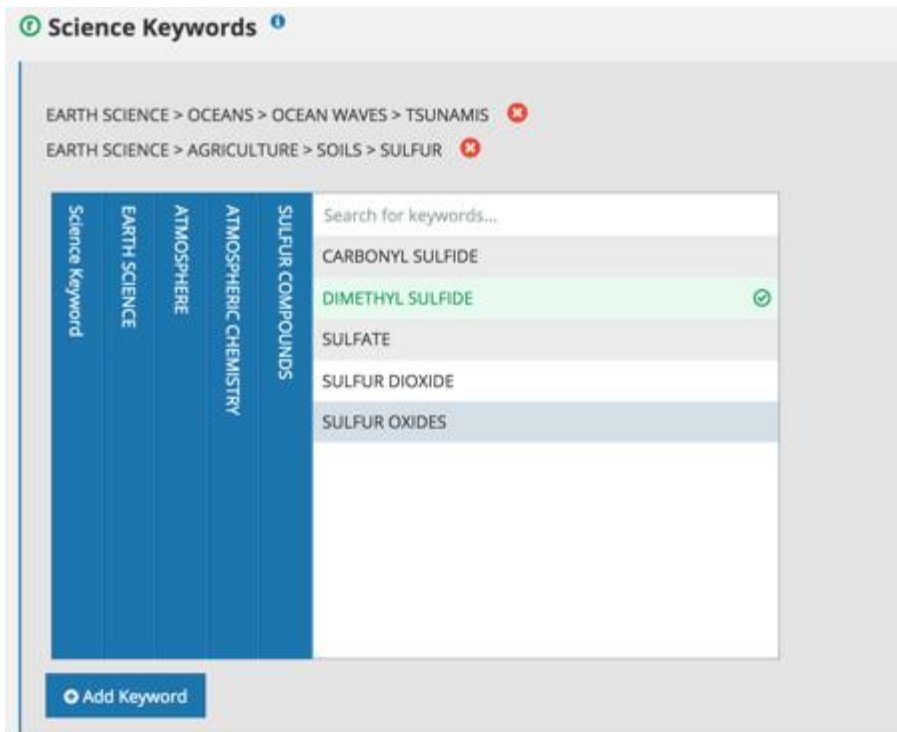
- Required but not complete
- Required and complete
- Optional and not complete
- Optional and complete
- Does not pass validation

Close

# CMR Ingest Validation



# MMT and CMR Keyword Validation



```
curl -i -XPOST ... -H "Cmr-Validate-
Keywords: true" \
https://cmr.earthdata.nasa.gov/ingest/pro
viders/LARC_ASDC/validate/collection/some
NativeId -d \
```

```
"<Collection>
...
<ScienceKeywords>
  <ScienceKeyword>
    <CategoryKeyword>EARTH
SCIENCE</CategoryKeyword>

    <TopicKeyword>BIOOSPHERE</TopicKeyword>
      <TermKeyword>SOILS</TermKeyword>
    </ScienceKeyword>
  </ScienceKeywords>
</Collection>"
```

HTTP/1.1 422 Unprocessable Entity

```
<path>ScienceKeywords/0</path>
<error> Science keyword Category [EARTH
SCIENCE], Topic [BIOOSPHERE], and Term
```

# Preview Panel / Collection HTML



## Abstract

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) was developed jointly by the U.S. National Aeronautics and Space Administration (NASA) and Japan's Ministry of Economy, Trade, and Industry (METI). ASTER is capable of collecting in-track stereo using nadir- and aft-looking near infrared cameras. Since 2001, these stereo pairs have been used to produce single-scene (60- x 60-kilometer (km)) digital elevation models (DEM) having vertical (root-mean-squared-error) accuracies generally between 10- and 25-meters (m). The methodology used by Japan's Sensor Information Laboratory Corporation (SILC) to produce the ASTER GDEM involves automated processing of the entire ASTER Level-1A archive. Stereo-correlation is used to produce over one million individual scene-based ASTER DEMs, to which cloud masking is applied to remove cloudy pixels. All cloud-screened DEMs are stacked and residual bad values and outliers are removed. Selected data are averaged to create final pixel values, and residual anomalies are corrected before partitioning the data into 1 degree (°) x 1° tiles. The ASTER GDEM covers land surfaces between 83°N and 83°S and is comprised of 22,702 tiles. Tiles that contain at least 0.01% land area are included. The ASTER GDEM is distributed as Geographic Tagged Image File Format (GeoTIFF) files with geographic coordinates (latitude, longitude). The data are posted on a 1 arc-second (approximately 30-m at the equator) grid and referenced to the 1984 World Geodetic System (WGS84)/ 1996 Earth Gravitational Model (EGM96) geoid. While the ASTER GDEM 2 benefits from substantial improvements over GDEM 1, users are nonetheless advised that the products still may contain anomalies and artifacts that will reduce its usability for certain applications, because they can introduce large elevation errors on local scales. The data are provided "as is" and neither NASA nor METI/ERSDAC will be responsible for any damages resulting from use of the data. V002 data set release date: 2009-06-28 Data Set Characteristics: Geographic Extent: Global between 83° latitude Scene Coverage: 1° x 1° tiles Image Dimensions: 3601 x 3601 Total Number Tiles: V001: 22,604; V002: 22,702 Tile Volume: ~25MB, 6.4 MB compressed Compression Type: zip File Format: GEOTIFF Map Projection: Geographic Lat/Lon Datum: WGS84/EGM96 Resolution: 1 arcsecond (30-m horizontal posting at equator)



### Spatial Coordinates

**Bounding Rectangle**

N: 82.0 S: -83.0 E: 180.0 W: -180.0

### Location Keywords

No Spatial Keywords found

### Temporal Coverages

**DateTime Ranges**

1999-12-18 to 2011-02-28

# Help Text

The screenshot shows a web browser window titled "Metadata Management Tool" with the URL [https://mmt.uat.earthdata.nasa.gov/drafts/147/edit/acquisition\\_information](https://mmt.uat.earthdata.nasa.gov/drafts/147/edit/acquisition_information). The main interface is for editing "Platform 1". It includes fields for "Type" (set to "Earth Observation Satellites"), "ShortName" (set to "terra"), and "LongName". A "Characteristics" section is visible, containing a table with columns "Name", "Unit", and "Data Type". A help dialog box is open, titled "Characteristics", providing instructions on platform-specific characteristics and their validation rules.

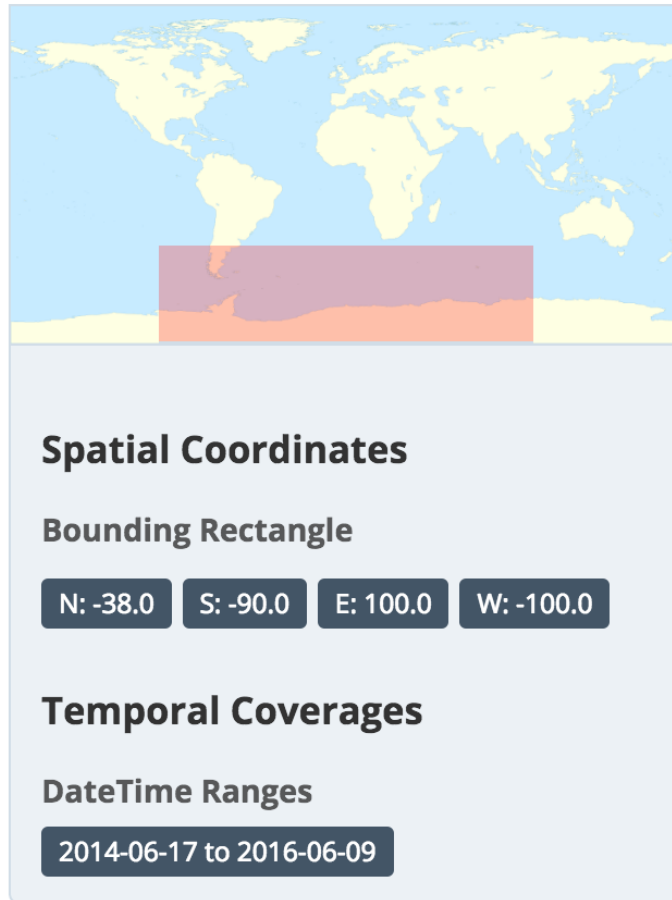
**Characteristics**

Platform-specific characteristics, e.g., Equator Crossing Time, Inclination Angle, Orbital Period. The characteristic names must be unique on this platform; however the names do not have to be unique across platforms.

**Validation**

- Minimum Items: 0

Close



# **METADATA QUALITY RUBRICS**

# Metadata Quality Rubrics

- Provide indication of completeness of Metadata based on different recommendations (UMM-C, CSW, DataCite...)
- Automatically evaluated and tagged in CMR
- Badges and guidance displayed in MMT.
  - Encourages better quality metadata
- Future
  - Reports, aggregations, search relevancy.


Metadata Management Tool

Mark

← → ↺

https://mmt.uat.earthdata.nasa.gov/drafts/147

☆ 🛑 ❤️ 💬 📁 🗺️ 🗣️ 📷 🌐 🌐 ☰

EARTHDATA

Metadata Management Tool

BETA

MMT\_1

Mark Reese Change Provider Logout

Quick Find

Enter Search Term

Find

Full Metadata Record Search

Manage Metadata Drafts Test-Record\_001

Test-Record\_001

Kathy's 101 test record - Cloned

DRAFT RECORD

VERSION 001

☆☆☆☆☆

Quality Score: 20

Required fields not complete

Publish Draft Delete Draft

Metadata Fields ⓘ

Collection Information

○

r

r

r

○

○

○

Descriptive Keywords

✓

r

○

○

Spatial Information

✓

r

○

○

○

Collection Citations

✓

○

Data Identification

✓

r

○

r

○

○

○

○

○

○

Acquisition Information

✓

r

○

Organizations

○

○

○

Metadata Information

✓

○

○

○

Distribution Information

✓

r

○

Temporal Information

✓

r


○

○

Personnel

✓

○

 EOSDIS

18

SESIP-0716-JG

# HUMANIZERS

# Humanizers Prototype

- Give clients clean, consistent facet data without having to change the underlying metadata.
- Temporary solution to help users while metadata is cleaned up.
- Humanizers can inform changes that are needed to metadata.

# Fixing Problems in Facets

- Misspellings: “Bioosphere”
- Legacy Terms: “AM-1” instead of Terra
- Inconsistent Names: Processing levels “Level 1”, “1”
- Whitespace around element values
- Use normal case when appropriate

# Before Humanizers

The image displays two screenshots of the NASA EarthData Search interface, illustrating the search results for 'Terra' and 'AM-1' before humanization.

**Top Screenshot: Search for 'Terra'**

- Search Bar:** Terra
- Results:** 630 Matching Collections
- Browse Collections:** SPOT-4 (17), SPOT-5 (15), **TERRA (630)**, THEOS (12), UK-DMC (12), Instrument
- Recent and Featured:** ASTER Global Digital Elevation Model V002 (ASTGTM v002 - LPDAAC, 1999-12-18 to 2011-02-28, 22702 Granules)

**Bottom Screenshot: Search for 'AM-1'**

- Search Bar:** AM-1
- Results:** 48 Matching Collections
- Browse Collections:** Keywords, Project, Processing level, Organization, Platform, **AM-1 (48)**, Instrument
- Recent and Featured:** ASTER L1A Reconstructed Unprocessed Instrument Data V003 (AST\_L1A v003 - LPDAAC, 1999-12-18 ongoing, 2823846 Granules)

# After Humanizers

The screenshot shows the NASA EarthData Search website. The top navigation bar includes links for Data Discovery, DAACs, Community, and Science Disciplines. The search bar contains the text 'Terra'. Below the search bar, a sidebar on the left lists 'Browse Collections' with a table of satellite instruments and their collection counts. The main content area displays '630 Matching Collections' and a 'Recent and Featured' section highlighting the 'ASTER Global Digital Elevation Model V002'.

Browse Collections	
SPOT-1	10
SPOT-4	17
SPOT-5	15
<b>TERRA</b>	<b>x678</b>
THEOS	12
UK-DMC	12

**630 Matching Collections**  
Add collections to your project to compare and retrieve their data.  
Search Time: 1.1s  
[Report a metadata problem](#)

**Recent and Featured**

**ASTER Global Digital Elevation Model V002**  
ASTGTM v002 - LPDAAC  
1999-12-18 to 2011-02-28 | 22702  
Granules

v 1.19.3 • NASA Official: Andrew Mitchell • FOIA • NASA Privacy Policy • USA.gov

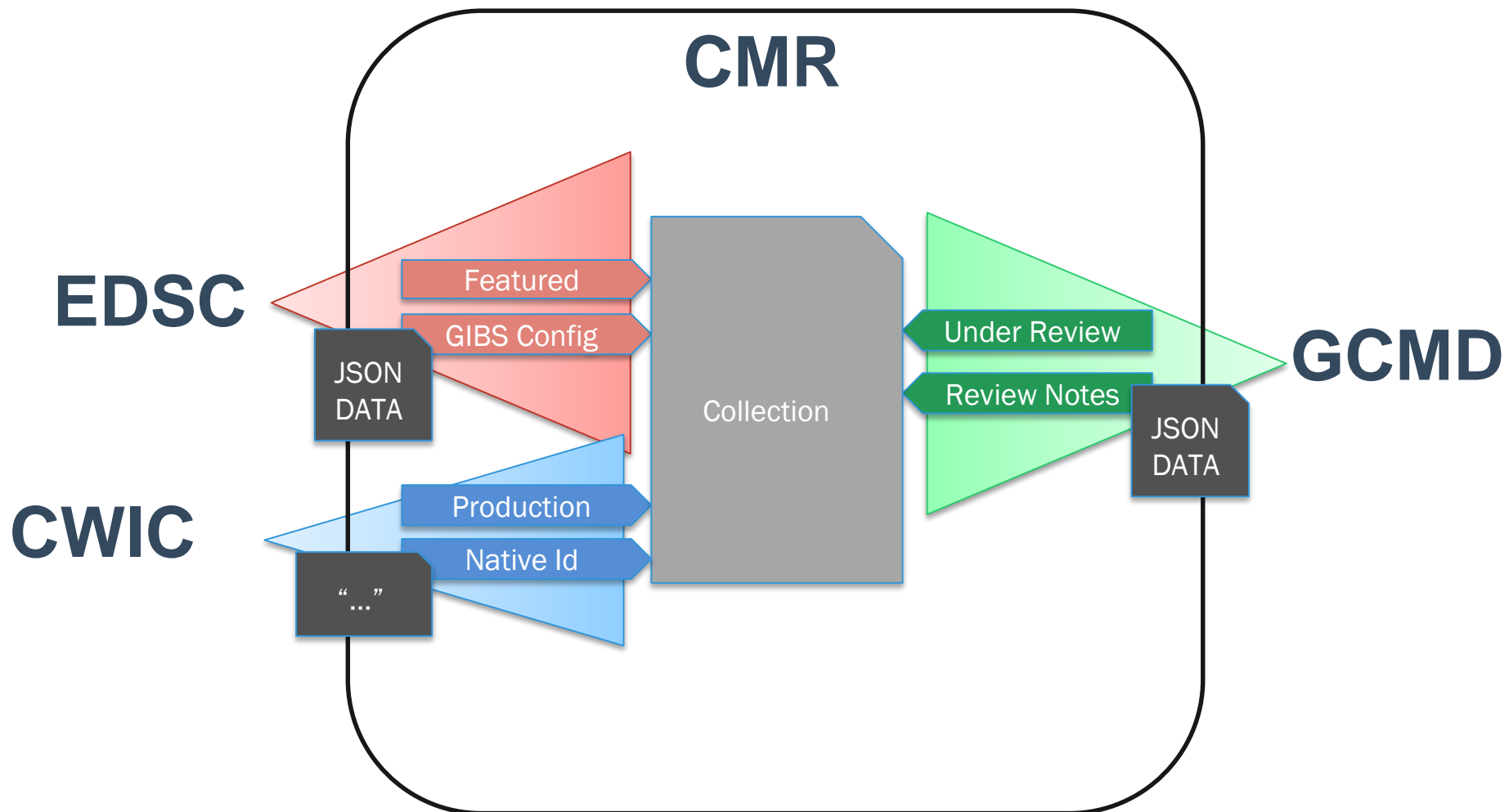
# How it works

1. Admin creates humanizer instructions in CMR
2. CMR indexes impacted collections with additional humanized fields
3. Clients optionally request facets with humanized values.

Underlying metadata and existing fields are **not modified**.

# TAGS

Tags  
=  
Enhanced Metadata



# What is a Tag?

- Key
  - “edsc.extra.gibs”
- Description
  - “Used to associate GIBS configuration for a collection”
- Category (optional)
  - “EDSC\_display”

# What is a Tag Association?

- Connects a Tag with a Collection
- Tag Key
  - “edsc.extra.gibs”
- Collection Id
  - “C1000000560-NSIDC\_ECS”
- Data
  - Any arbitrary JSON up to 32K. Single strings will be made searchable.

# Tag Data can be Retrieved

```
curl "https://cmr.../search/collections.json?\n      tag_key=edsc.extra.gibs\n      &include_tags=edsc.extra.gibs"
```

```
...
"tags" : {
  "edsc.extra.gibs" : {
    "data" : [ {
      "maxNativeZoom" : 5,
      "antarctic" : false,
      "product" : "GHRSSST_L4_G1SST_Sea_Surface_Temperature",
      "geo" : true,
      "arctic" : false,
      "title" : "Sea Surface Temperature (L4, G1SST)",
      "source" : "Multi-mission / GHRSSST",
      "match" : {
        "time_start" : ">=2010-06-21"
      },
      "format" : "png",
      "resolution" : "1km"
    } ]
  }
}
```

# Tag Data can be Searched

Find all collections:

- tagged with “org.ceos.wgiss.cwic.granules.native\_id”
- with data “JPL-L2P-MODIS\_A”

```
curl "https://cmr.../search/collections.json?\
```

```
tag_data[org.ceos.wgiss.cwic.granules.native_id]=JPL-L2P-MODIS_A"
```

# Tags Enable Many Use Cases

- Allows layering on additional information to each collection.
- Examples
  - Mark “featured” collections.
  - Categorize collections (Reviewed, approved, needs work, etc).
  - Store visualization information
  - Add a client specific id to collections.

This material is based upon work  
supported by the National  
Aeronautics and Space  
Administration under Contract  
Number **NNG15HZ39C.**

**Raytheon**